Infrastructure Provided by Government Team Members

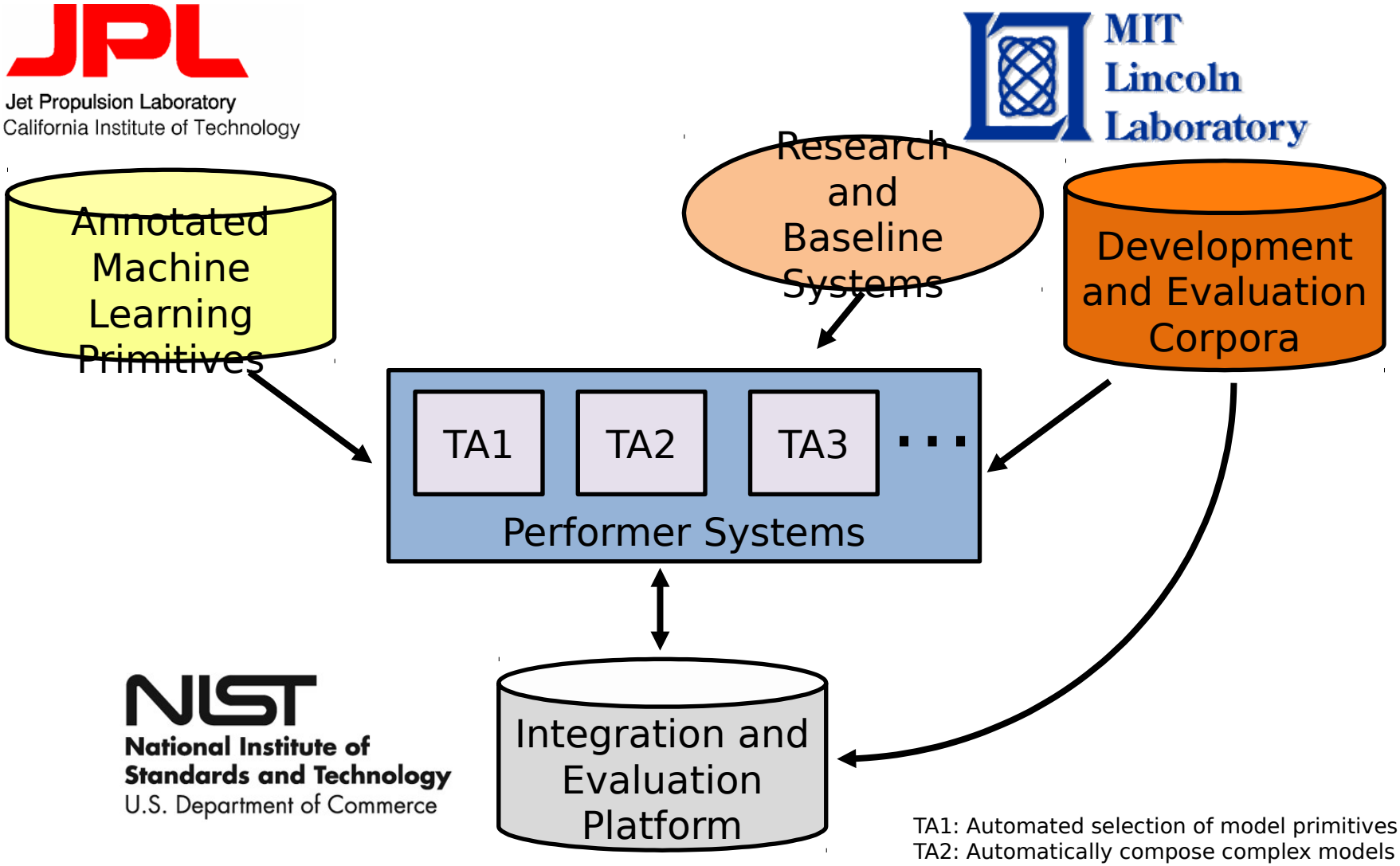
Chris Mattmann - Jet Propulsion Laboratory
Richard Lippmann - MIT Lincoln Laboratory
Mark Przybocki – National Institute of Standards and Technology

DARPA D³M

Proposer's Day

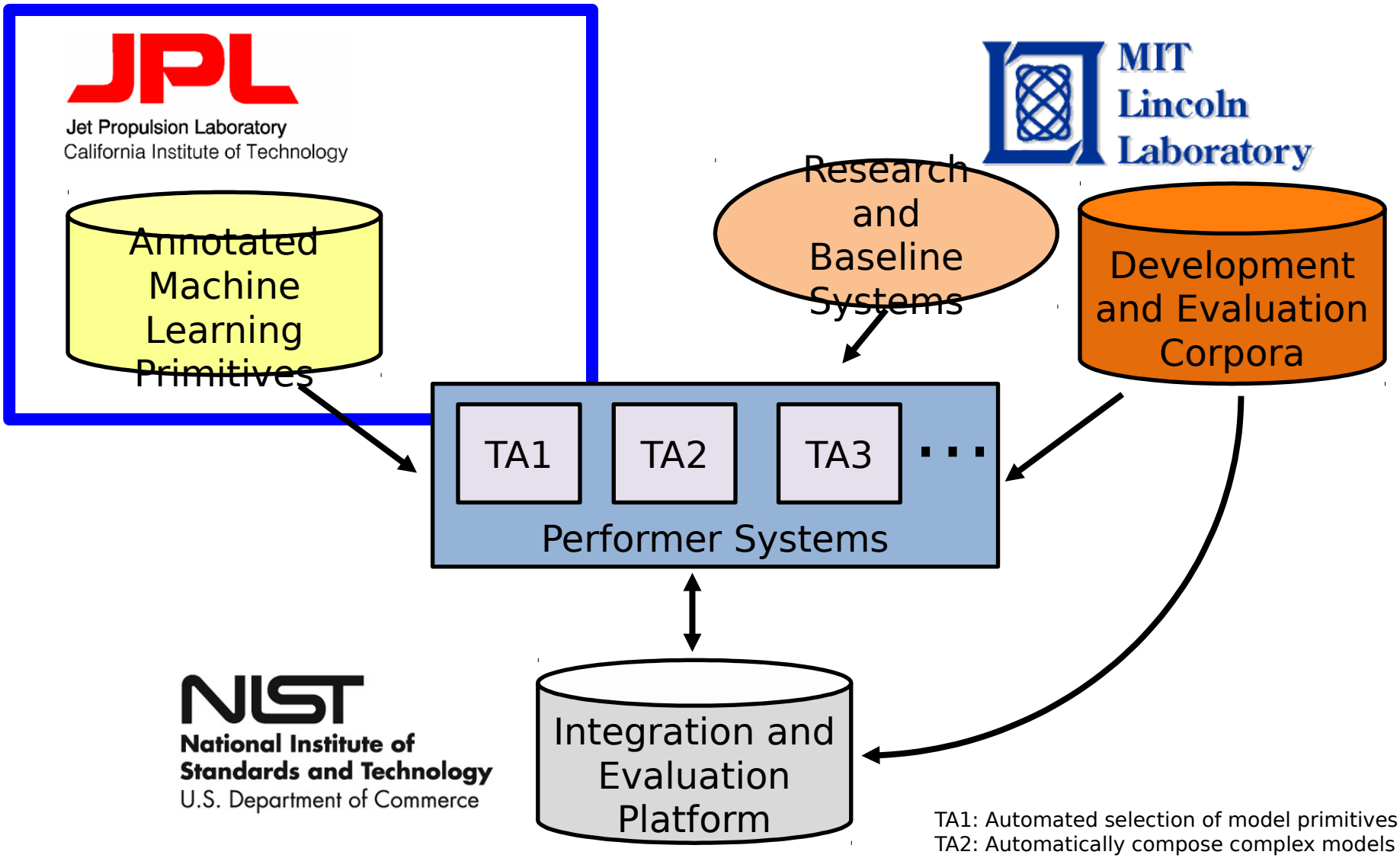
14 June 2016

Infrastructure Provided by Government Team Members



TA1: Automated selection of model primitives
TA2: Automatically compose complex models
TA3: Model curation by non-expert users

Jet Propulsion Laboratory Roles



TA1: Automated selection of model primitives
TA2: Automatically compose complex models
TA3: Model curation by non-expert users

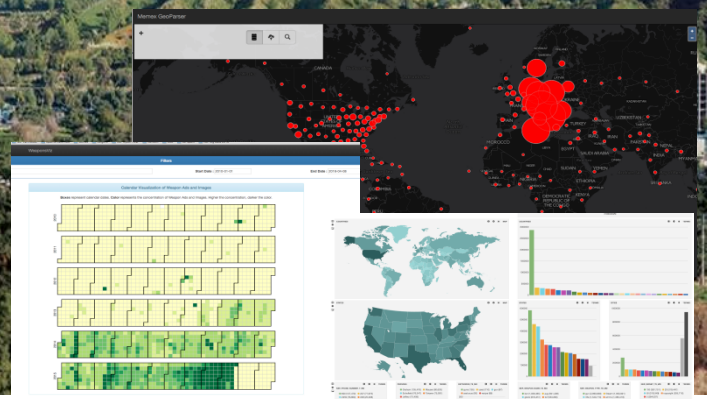
NASA Jet Propulsion Laboratory (JPL) / Caltech

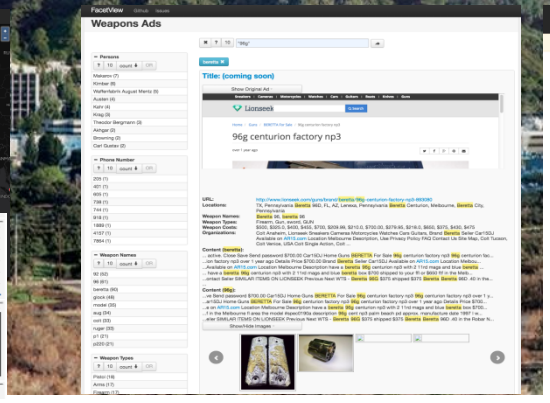
- JPL is a child of Caltech: founded in 1936 as a graduate student project under Professor Theodore von Kármán.
- JPL led the development of US rocket technology in WWII. Caltech and JPL staff founded Aerojet Corporation, the first US rocket firm.
- JPL worked in collaboration with Werner von Braun and German rocket engineers to create the US missile program from 1946 to 1958.
  - JPL worked under a US Army Ballistic Missile Agency contract to design and build the Corporal and Sergeant, the first US ballistic missiles.
- After Sputnik, JPL was transferred to NASA upon its creation in 1958.

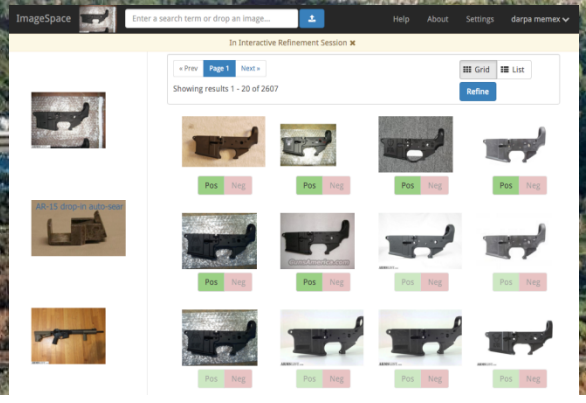
Jet Propulsion Laboratory
California Institute of Technology

Has a dual characte
A unit of Caltech, staffed with Caltech employees;
A Federally-Funded Research and Development Center (FFRDC) under NASA sponsorship;
Is a major national research and development (R&D) capability supporting:
NASA programs;
Defense programs;
Civil programs of national importance compatible with JPL capabilities.

DARPA involvement in XDATA and Memex initiatives building Apache open source information retrieval – 2012 - present

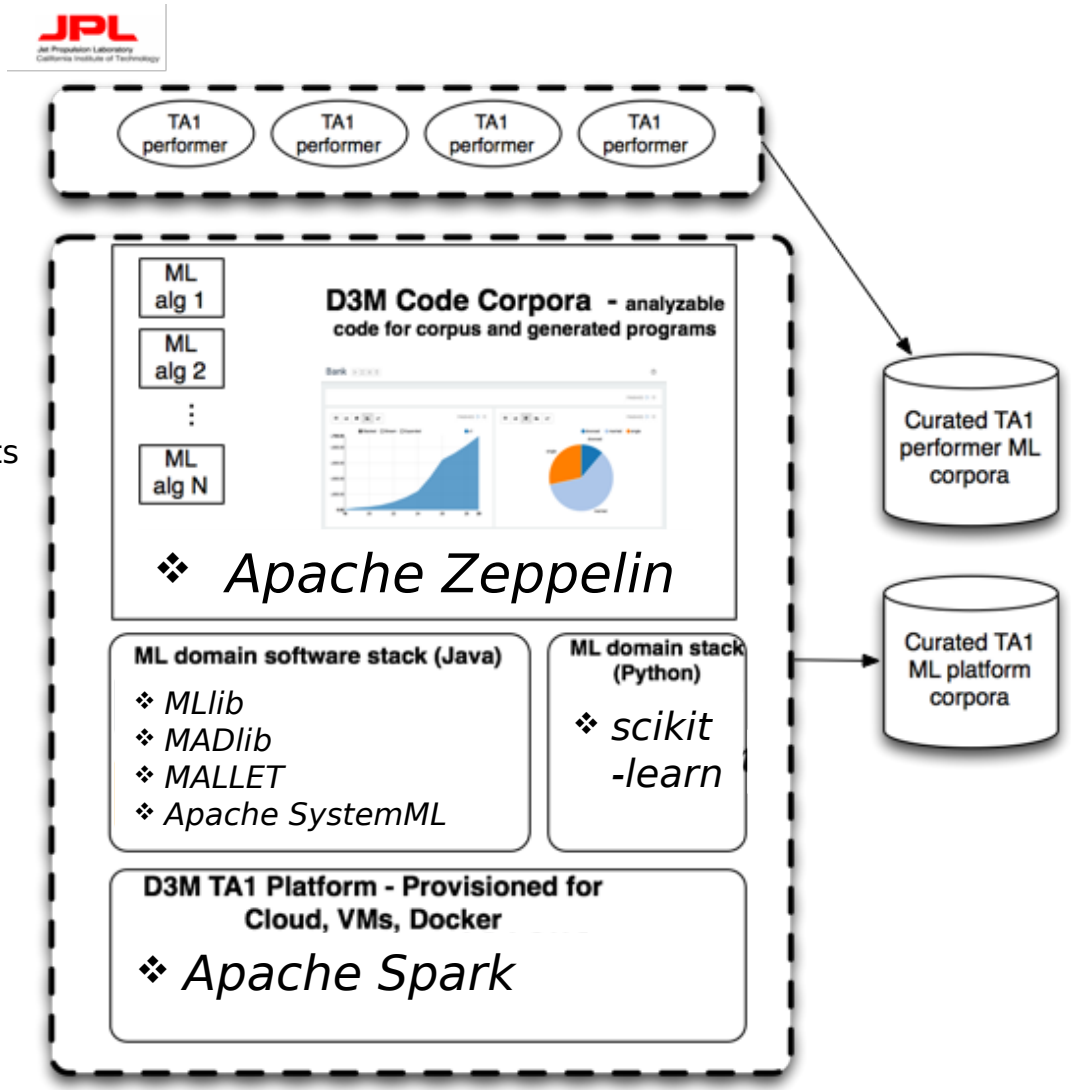






TA1 Framework – ML primitives and infra

- TA1 platform supporting
  - Python, Java/Scala ecosystem (core)
  - Support for C++/C, R, Julia
- Focused on permissive and open source tech
  - Apache Spark as a scaling platform
  - ML libraries and frameworks that run on
    Spark including new(ish) Apache efforts
  - Scikit Learn and Apache Spark
- Work with MIT-LL and TA1 performers
  to annotate ML code for composability
  - Focus on notebooks and evaluating Apache Zeppelin
- Generate analyzable ML code corpus
  from TA1 and jointly annotate with
  MIT-LL and work with to NIST for evaluation
- Work with NIST to develop both a test/development cloud



Annotation Machine Learning Components for TA1 (Jointly with MIT Lincoln Laboratory)

Capabilities

- Classification
- Regression
- Feature Selection
- Extract Text/Documents from Web Pages
- Cluster Documents
- Calibrate posterior probabilities

Parameters

Component	**Parameters**	
	Discrete	Continuous
AdaBoost	1	3
Linear SVM	2	2
Multi-Layer Perceptron	8	15
Random Forest	2	3
Normalize	1	

Inputs/Outputs

- Can handle: numeric, categorical, string, relational inputs
- Can handle: varying numbers of input features
- Can handle: missing input features
- Provides class likelihood probabilities
- Provides class posterior probabilities

Infrastructure Thoughts

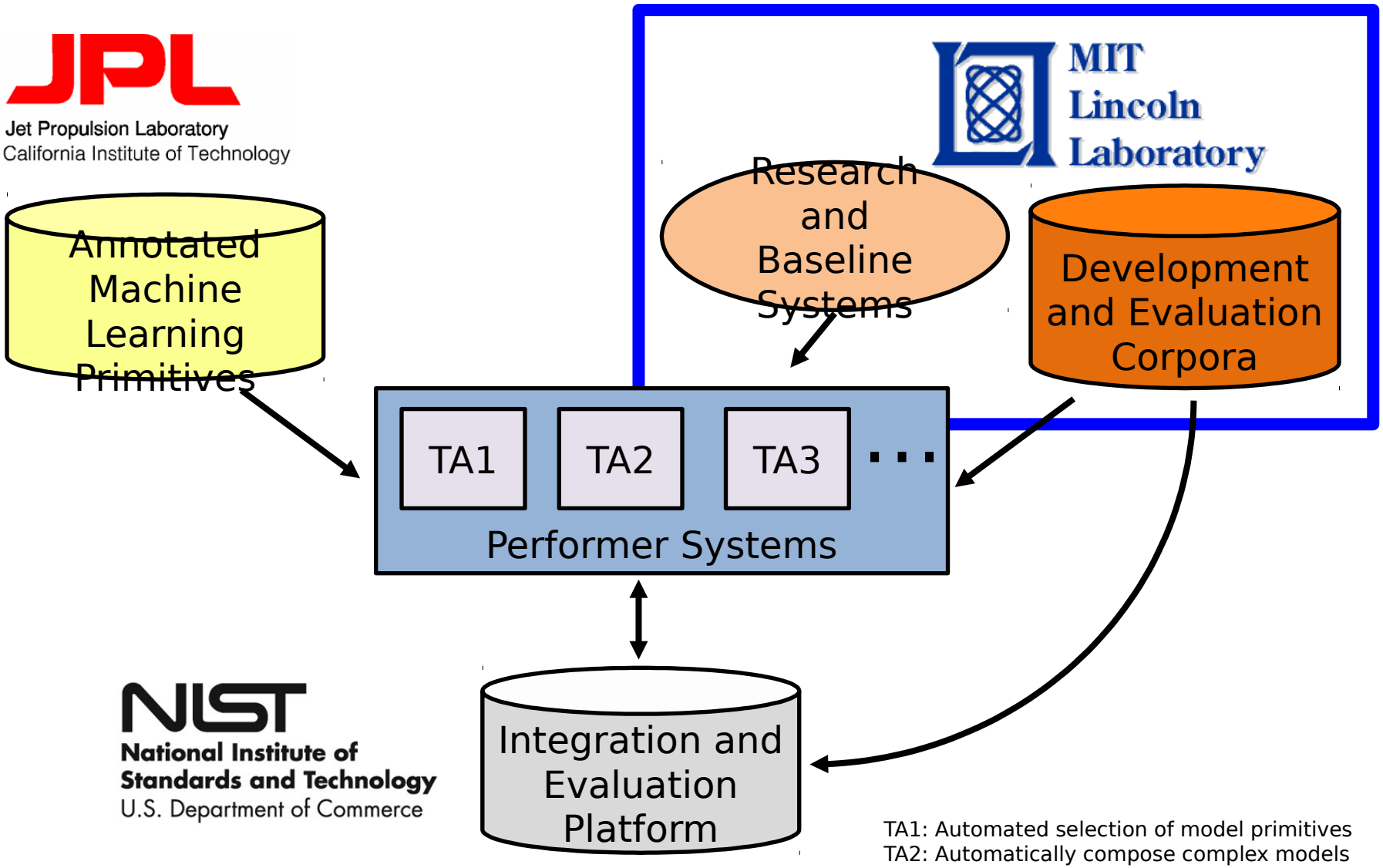
❖ *Docker*

- On-premise and cloud support
- Will rely on container technologies including Docker
- Willingness to incorporate useful components and processes
  - Goal is to be up and running with a corpus and library when the program starts
- Infrastructure will include two environments
  - An “algorithm testbed” and corpus – will be up and running and used throughout the program for integration and test – stood up by JPL in coordination with NIST
  - An “evaluation testbed” – used by NIST to evaluate the performers ML components and processes

Some Open Questions and Thoughts for TA1

- Use existing APIs to take advantage of capabilities out there
  - MeaningCloud API as a potential for ground-truth – other possibilities include ICE: http://research.microsoft.com/apps/pubs/default.aspx?id=230042
  - Amazon Mechanical Turk and intelligent survey/system design
- Considerations if you're using our provided infrastructure during I&T
  - We're going to providing all these ML components and we'd love for you to use them
    1. *Docker-ized containers*
    2. *Continuous Integration*
    3. *Will need to support an algorithm dev testbed (JPL) and NIST eval hardware*
- Some considerations if you are rolling your own infrastructure for I&T
  - GPU-focused code and algorithms and abstraction
    1. *Requiring specialized hardware is possible if it offers order of magnitude speed-ups; tradeoff there between running on commodity over e.g., Apache Spark – should address this in proposed work*
  - TA1 / TA2 integration lines
    1. *Open models, open data, open source (ALv2, BSD, MIT, permissive)*
- Note: for evaluation you will all use the same NIST infrastructure
  - Should be a copy of the I&T environment

MIT Lincoln Laboratory Roles



TA1: Automated selection of model primitives
TA2: Automatically compose complex models
TA3: Model curation by non-expert users

MIT Lincoln Laboratory Department of Defense (DoD) Federally Funded R&D Center

Massachusetts Institute of Technology

MIT Lincoln Laboratory, Lexington, Massachusetts

Mission: Technology in Support of National Security

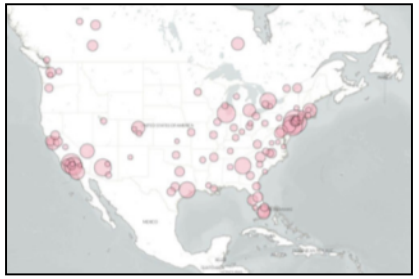
Key Roles: System architecture engineering
Long-term technology development
System prototyping and demonstration

Big Data Toolkit and Applications (XDATA, Memex)

Structured, Semi-Structured, Unstructured Data

- *VizLinc*
- *TweeT Tools*
- *LLClass*
- *MITIE*

Extraction Technologies for Text, Speech, Images, Cyber



Structured Data, Analytics, User Interface

Task 1. Select Problems and Corpora for System Development and Evaluations

MIT Lincoln Laboratory

In close collaboration with NIST and JPL

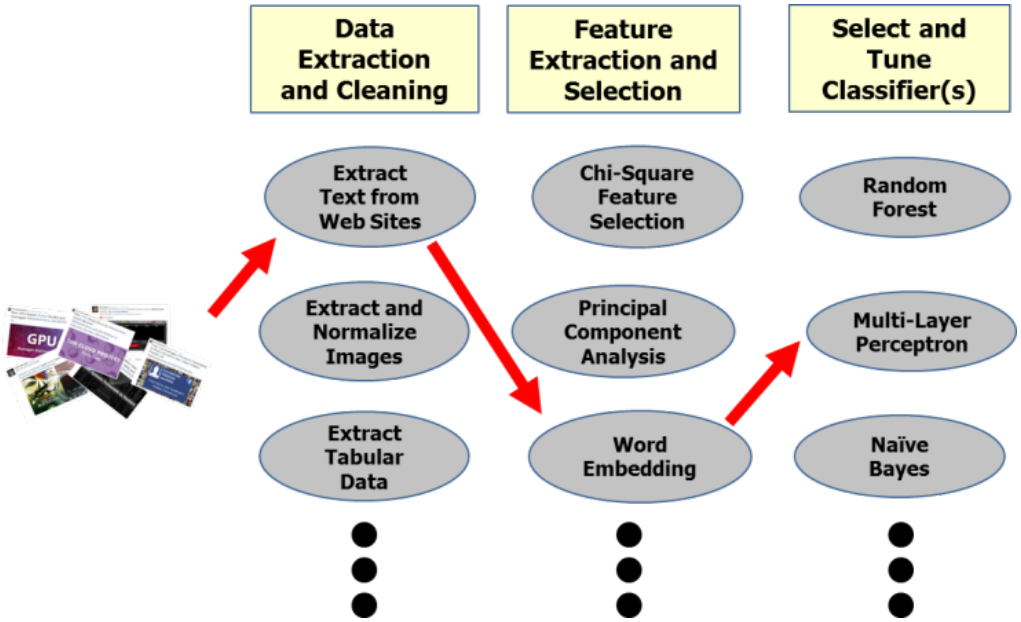
- *Open Science Data Cloud Partnerships for International Research and Education (OSDC PIRE)*
- *Kaggle*
- *Dataverse*
- *UCI*

- Problems and data will be open but have characteristics similar to a range of problems of interest
- Systems must work from raw input data
- Performance requirements will come from subject matter experts in the application domains
- Subject matter experts will be available for TA-3 evaluations requiring user-in-the loop testing

MIT Lincoln Laboratory

- Include classification, detection, and regression problems
- Left-out labeled data will be used for yearly blind evaluations
- Initially, the machine learning problems will be simpler, but they will gradually become more difficult due to …
  - Fewer provided labels with the development data
  - Less cleaned raw data that is noisier and more difficult to automatically analyze
  - Imbalanced class prior probabilities
- The amount of data provided will be that actually available in selected problem domains

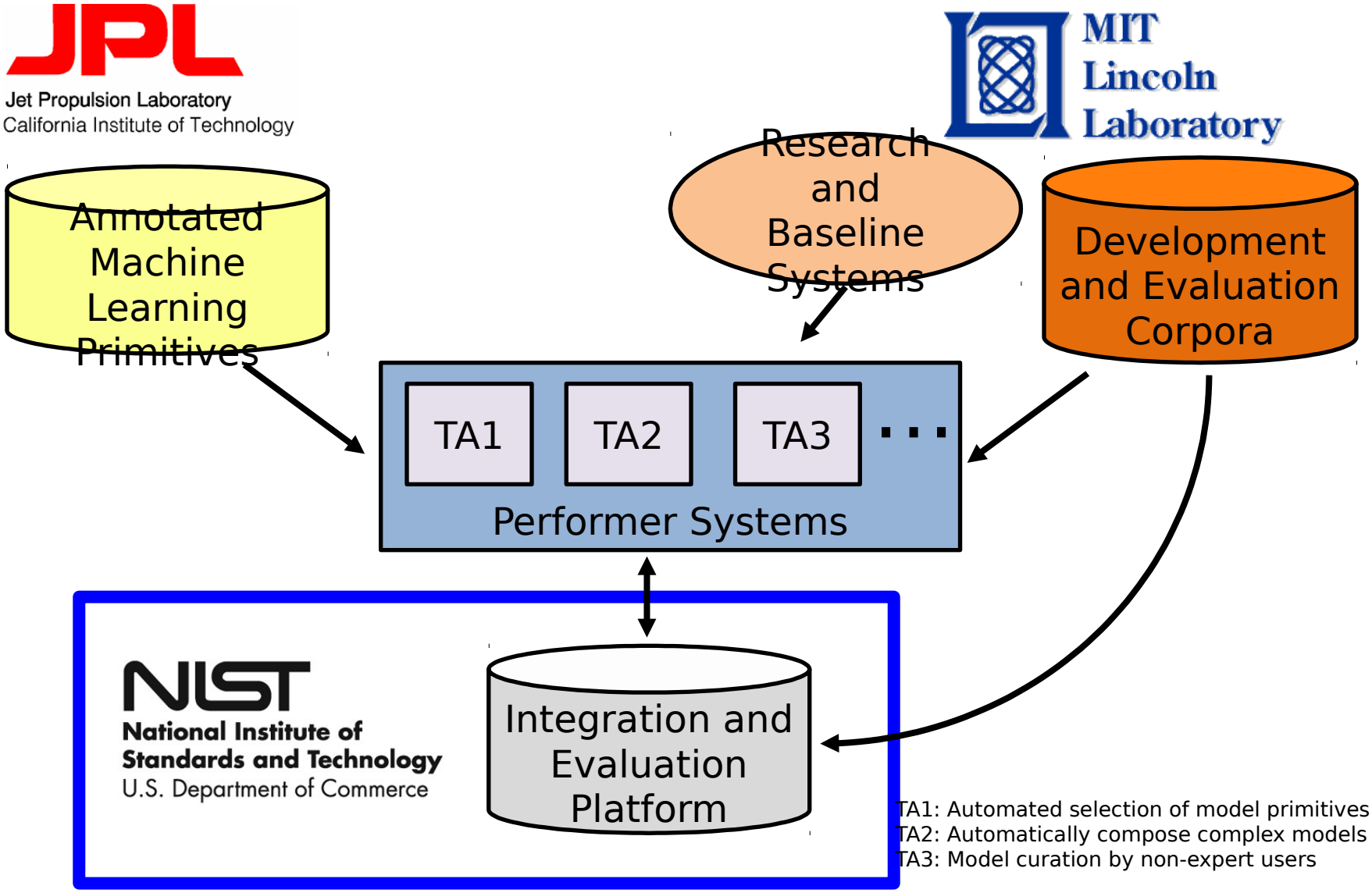
Task 2. Research and Baseline Automated Systems



Example Components in a TA-2 System

- Create automated machine learning systems for TA-2 and TA-3
- Validate machine learning primitives and their annotation
- Assess the difficulty of problems and corpora
- Provide baseline scores and explore new approaches
- Open-source systems to simplify entry to evaluations

NIST Roles


JPL
Jet Propulsion Laboratory
California Institute of Technology
MIT
Lincoln
Laboratory
Annotated
Machine
Learning
Primitives
Research
and
Baseline
Systems
Development
and Evaluation
Corpora
TA1
TA2
TA3
. . .
Performer Systems
NIST
National Institute of
Standards and Technology
U.S. Department of Commerce
Integration and
Evaluation
Platform
TA1: Automated selection of model primitives
TA2: Automatically compose complex models
TA3: Model curation by non-expert users

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

Multimodal Information Group (MIG)

www.nist.gov/itl/iad/mig

NIST	… promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life …
MIG (Group)	… research and develop measurement and evaluation methods to advance technologies that provide more effective access to multimedia and multi-lingual information… *speech – text – images – videos – multimodal data – media fusion* ● Speech Analytics ● Image & Video Analytics ● Machine Translation ● Speaker and Language Recognition ● Data
MIG Data Scienc e	… supports a larger NIST effort in Big Data by advancing test, evaluation, and benchmarking methods for emerging data analytics … ● DSE: (new) Data Science Evaluation Series – (TREC model) yearly evaluation tracks for generalizable data analytic problems ● EMS: (in house) Evaluation Management System – private

Role in D3M Integration and Evaluation Platform

National Institute of Standards and Technology
U.S. Department of Commerce

In close collaboration with JPL and MIT-LL

Infrastructure

- Define API's
  - ML Module upload
  - System piping of modules
  - Data access
- Develop T&E support SW
- Online scoring server

Evaluation(s)

- Evaluation Plan
  - tasks, data, rules, schedule, …
- Yearly "open" Evaluations
- Collaborate on Data
- Publish Results
- Evaluation Workshop

Measurement/Metrics

- Define Human-in-Loop
  - SME vs Naïve user (time, accuracy, …)
- Accuracy
  - Relative to baseline systems
- Resource Benchmarking
- Ground-truth

Demonstration

- Extend Evaluation Platform (harness) to "demo" systems

Measurement and **Evaluation techniques** will vary across TA1, TA2, TA3 and program level evaluations

Evaluation metrics

Develop and publish protocols and definitions for:

TA1: Measure efficiency of predicting primitives

Protocol	Metric
Predict primitives; experts compose	Δ error of predicted primitives vs. optimal

TA2: Measure (re)discovery of optimal analyses/models

Protocol	Metric
Synthesize models compare to experts	Δ error of D³M model vs. expert models

TA3: Human curation of models

Protocol	Metric
Decompose questions, compare with experts	P_d/P_{fa} of automated vs. expert (ROC)
Compare decisions made by experts with lay users	Δ error of analyst vs. data scientist in-the-loop

Program-level evaluation (annual integration and evaluation)

Protocol	Metric
TA1-3 form team, work with non-experts to build model, compare with experts	Δ error of D³M model vs. expert model